

## **Cayley-Encoding Of Unitary Matrices For Differential Communication**

### **Field of the Invention**

[0100] The present invention is directed toward a method of communication in which the input data is encoded using unitary matrices, transmitted via one or more antennas and decoded differentially at the receiver without knowing the channel characteristics at the receiver, and more particularly to technology for producing a set of unitary matrices for such encoding/decoding using Cayley codes.

### **Continuing Application Information**

[0101] The present application is a Continuation-In-Part ("CIP") under 35 U.S.C. § 120 of copending U.S. Patent Application Serial No.09/356,387 filed July 16, 1999, the entirety of which is hereby incorporated by reference.

[0102] The present application also claims priority under 35 U.S.C. § 119(e) upon Provisional U.S. Patent Application Serial No. 60/269,838 filed February 20, 2001, the entirety of which is hereby incorporated by reference.

### **Background of the Invention**

[0103] Although reliable mobile wireless transmission of video, data, and speech at high rates to many users will be an important part of future telecommunications systems, there is considerable uncertainty as to what technologies will achieve this goal. One way to get high rates on a scattering-rich wireless channel is to use multiple transmit and/or receive antennas. Many of the practical schemes that achieve these high rates require the propagation environment or channel to be known to the receiver.

[0104] In practice, knowledge of the channel is often obtained via training: known signals are periodically transmitted for the receiver to learn the channel, and the channel parameters are tracked (using decision feedback or automatic-gain-control (AGC)) in between the transmission of the training signals. However, it is not always feasible or advantageous to use training-based schemes, especially when many antennas are used or either end of the link is moving so fast that the channel is changing very rapidly.

[0105] Hence, there is much interest in space-time transmission schemes that do not require either the transmitter or receiver to know the channel. A standard method used to combat fading in single-antenna wireless channels is differential phase-shift keying (DPSK). In DPSK, the transmitted signals are unit-modulus (typically chosen

from a  $m$ -PSK set), information is encoded differentially on the phase of the transmitted signal, and as long as the phase of the channel remains approximately constant over two consecutive channel uses, the receiver can decode the data without having to know the channel coefficient.

[0106] Differential techniques for multi-antenna communications have been proposed, where, as long as the channel is approximately constant in consecutive uses, the receiver can decode the data without having to know the channel. The general differential techniques of the Background Art have good performance when the set of matrices used for transmission forms a group under matrix multiplication, which also leads to simple decoding rules.

[0107] But the number of groups available is rather limited, and the groups do not lend themselves to very high rates (such as tens of bits/sec/Hz) with many antennas. One of the Background Art techniques is based on orthogonal designs, and therefore has simple encoding/decoding and works well when there are two transmit and one receive antenna, but suffers otherwise from performance penalties at very high rates.

[0108] Part of the difficulty of designing large sets of unitary matrices is the lack of simple parameterizations of these matrices. To keep the transmitter and receiver complexity low in multiple antenna systems, linear processing is often preferred, whereas unitary matrices are often highly nonlinear in their parameters.

### **Summary of the Invention**

[0109] The invention, in part, provides a partial and yet robust solution to the general design problem for differential transmission, for rate  $R$  (in bits/channel use) with  $M$  transmit antennas for an unknown channel.

[0110] The invention, also in part, provides Cayley Differential ("CD") codes that break the data stream into substreams, but instead of transmitting these substreams directly, these substreams are used to parameterize the unitary matrices that are transmitted. The codes work with any number of transmit and receive antennas and at any rate. The Cayley code advantages include that they:

1. Are very simple to encode;
2. Can be used for any number of transmit and receive antennas;
3. Can be decoded in a variety of ways including simple polynomial-time linear-algebraic techniques such as: (a) Successive nulling and canceling and (b) Sphere decoding;
4. Are designed with the numbers of both the transmit and receive antennas in mind; and

5. Satisfy a probabilistic criterion namely, maximization of an expected distance between matrix pairs.

[0111] Additional features and advantages of the invention will be more fully apparent from the following detailed description of the preferred embodiments, the appended claims and the accompanying drawings.

#### **Brief Description of the Drawings**

[0112] Figure 1 is a plot of the performance of an example CD code according to an embodiment of the invention for the example circumstances of  $M=2$  transmit and  $N=2$  receive antennas with rate  $R=6$ .

[0113] Figure 2 is a plot of block error performance of an example CD code according to an embodiment of the invention for the example circumstances of  $M=4$  transmit and  $N=1$  receive antennas with rate  $R=4$  compared with a nongroup code.

[0114] Figure 3 is a plot of the performance of an example CD code according to an embodiment of the invention for the example circumstances of  $M=4$  transmit and  $N=2$  receive antennas with rate  $R=4$ .

[0115] Figure 4 is a plot of the performance of an example CD code according to an embodiment of the invention for the example circumstances of  $M=4$  transmit and  $N=4$  receive antennas with rate  $R=8$ .

[0116] Figure 5 is a plot of the performance of an example CD code according to an embodiment of the invention for the example circumstances of  $M=8$  transmit and  $N=12$  receive antennas with rate  $R=16$ .

[0117] Figure 6 is a flow chart of steps included in a transmitting method according to an embodiment of the invention.

[0118] And Figure 7 is a flow chart of steps included in a receiving method according to an embodiment of the invention.

[0119] The accompanying drawings are: intended to depict example embodiments of the invention and should not be interpreted to limit the scope thereof; and not to be considered as drawn to scale unless explicitly noted.

#### **Detailed Description of Preferred Embodiments**

[0120] Using Cayley Differential ("CD") codes for differential unitary space-time (DUST) communication includes two aspects. The first is transmitting/receiving data with the CD codes, which necessarily involves calculating the CD codes. The second aspect is designing how the actual CD codes will be calculated. It should be noted that the first aspect assumes that the second aspect has been performed at least once.

**[0121]** The first aspect of Cayley-encoded DUST communication itself includes two aspects, namely (A) Cayley-encoding and transmitting, and (B) receiving and Cayley-decoding.

**[0122]** A few statements about notation will be made. The notation  $\alpha_1, \dots, \alpha_Q$  indicates a set of real-valued scalars each member of which is referred to as a parameter  $\alpha$ , there being a total of  $Q$  such parameters. Another way to refer to the set  $\alpha_1, \dots, \alpha_Q$  is via the notation:  $\{\alpha_q\}$ . Each  $\alpha_q$  will take its value to be one of the elements in the set known as  $A$  (the reader should be careful to note that this fancy symbol is the letter "A" in the Euclid Math One font in order to distinguish from a Hermitian basis matrix known as "A" which will be discussed below). Mention will also be made of a set of matrices referred to as  $\{A\}$ , as well as a subset having  $Q$  elements, referred to as  $\{A_q\}$ .

**[0123]** As depicted in Fig. 6, transmitting  $R \times M$  bits of Cayley-encoded data (starting at step 602) can include: (step 604) breaking the  $R \times M$  total bits of data to be transmitted into  $R \times M / Q$  chunks; (step 606) mapping each of the  $Q$  chunks of  $R \times M / Q$  bits to take one of the different scalar values found in the set  $A$ , where  $A$  has  $r = 2^{RM/Q}$  elements and  $A$  has been previously determined ( $A$  and its role will be discussed below), i.e., assigning a specific value from  $A$  to each chunk, respectively, in order to get  $\alpha_1, \dots, \alpha_Q$  (also referred to as " $\{\alpha_q\}$ ") (below  $\{\alpha_q\}$  and its role will be discussed); (step 608) calculating the specific CD code according to (Eq. No. 9, introduced below) using  $\{\alpha_q\}$  and  $\{A_q\}$  (where  $\{A_q\}$  has been previously determined; see the discussion below of the second aspect); (step 610) calculating the matrix to be transmitted (the "present matrix" representing the  $R \times M$  bits) based on the specific CD code and the previously-transmitted matrix according to the fundamental transmission equation (Eq. No. 4, introduced below); (step 612) modulating the present matrix on a carrier to form a carrier-level signal; and (step 616) transmitting the carrier-level signal (with flow ending at step 618).

**[0124]** As depicted in Fig. 7, receiving  $R \times M$  bits of Cayley-encoded data (starting at step 702) can include: (step 704) receiving a carrier-level signal; (step 706) demodulating the carrier-level signal to form a matrix (representing  $R \times M$  bits); (step 708) searching the set known as  $A$  to find  $Q$  specific scalar values, namely  $\alpha_1, \dots, \alpha_Q$  (or  $\{\alpha_q\}$ ), that minimize (Eq. No. 12, introduced below) or (Eq. No. 13, introduced below, which can be solved more quickly albeit less accurately than (Eq. No. 12)); (step 710) mapping each element of  $\{\alpha_q\}$  into its corresponding  $R \times M / Q$  bits using a

predetermined mapping relation; and (step 712) reassembling Q chunks of R\*M/Q bits to produce the R\*M bits of data that were transmitted (with flow ending at step 714).

**[0125]** The second aspect of Cayley-encoded DUST communication, namely the design of the CD codes, is to determine the non-data parameters of the CD code (Eq. No. 9, introduced below). This includes: determining the set of  $A_q$  (“ $\{A_q\}$ ”); and the  $A$  from which are selected  $\alpha_1, \dots, \alpha_Q$  (or  $\{\alpha_q\}$ ). The  $\{A_q\}$  is independent of the data to be transmitted but is dependent upon the transmitter/receiver hardware. Both  $\{A_q\}$  and  $A$  will be known to the transmitter and to the receiver.

**[0126]** Determining  $A$  can include: (1) substituting the number of transmitting antennas,  $M$ , and the number of receiving antennas,  $N$ , into (Eq. No. 15, introduced below) to get the number of degrees of freedom,  $Q$ ; (2) determining how many elements,  $r$ , will be in  $A$  according to the equation  $r = 2^{(RM/Q)}$  (a rewriting of (Eq. No. 22, introduced below)); (3) establishing the set of  $\theta$  (“ $\{\theta\}$ ”) according to the equation  $\{\theta\} = \{\pi/r, 3\pi/r, 5\pi/r, \dots, (2r-1)\pi/r\}$ ; and establishing  $A$  by substituting  $\{\theta\}$  into (Eq. No. 17, introduced below).

**[0127]** The  $\{A_q\}$  can be determined by solving (Eq. No. 18, introduced below) using the  $A$  determined above, e.g., via an iterative technique such as the gradient-ascent method. An iterative technique will find a  $\{A_q\}$  which causes (Eq. No. 18) to yield a maximum value (see (Eq. No. 19, introduced below)). But an iterative technique will not necessarily find the optimal  $\{A_q\}$  among all possible  $\{A\}$ , rather it will find the optimal value for that particular iterative technique being used.

**[0128]** Cayley codes, according to an embodiment of the invention are briefly summarized as follows. To generate a unitary matrix  $V$  parameterized by the transmitted data, we break the data stream into  $Q$  substreams (we specify  $Q$  later) and use these substreams to chose  $\alpha_1, \dots, \alpha_Q$  each from the set  $A$  with  $r$  real-valued elements (we also have more to say about this set later). A Cayley code of a rate  $R$ , where  $R = (Q/M) \log_2 r$  obeys the following equation (known as the Cayley Transform)

$$V = (I + iA)^{-1} (I - iA) \quad (1)$$

where

$$A = \sum_{q=1}^Q A_q \alpha_q \quad ,$$

$I$  is the identity matrix of the relevant dimension and  $A_1, \dots, A_Q$  are pre-selected  $M \times M$  complex Hermitian matrices.

[0129] The matrix  $V$ , as given by (2), is referred to as the Cayley transform of  $iA$  and is unitary.

[0130] The Cayley code is completely specified by  $A_1, \dots, A_Q$  and  $A$ . Each individual codeword is determined by the scalars  $\alpha_1, \dots, \alpha_Q$ .

[0131] The performance of a Cayley code according to an embodiment of the invention depends on the choices of the number of substreams  $Q$ , the Hermitian basis matrices  $\{A_q\}$ , and the set  $A$  from which each  $\alpha_q$  is chosen. Roughly speaking,  $Q$  is chosen so as to maximize the number of independent degrees of freedom observed at the output of the channel. To choose the  $\{A_q\}$ , a coding criterion can be optimized (to be discussed below) that resembles  $|\det(V_\ell - V_{\ell'})|$  but is more suitable for high rates and is amenable to analysis. The optimization need be done only once, during code design, and it is amenable to gradient-based methods. The Cayley transform (Eq. No. 1) is powerful because it generates the unitary matrix  $V$  from the Hermitian matrix  $A$ , and  $A$  is linear in the data  $\alpha_1, \dots, \alpha_Q$ .

[0132] For the purposes of the present application, it suffices to assume that the channel has a coherence interval (defined to be the number of samples at the sampling rate during which the channel is approximately constant) that is at least twice the number of transmit antennas.

### 1. Review of Differential Unitary Space-Time ("DUST") Modulation

[0133] The following is a brief summary of the differential, unitary matrix signaling scheme disclosed in the copending application, Serial No. 09/356,387, which has been incorporated by reference above in its entirety.

[0134] In a narrow-band, flat-fading, multi-antenna communication system with  $M$  transmit and  $N$  receive antennas, the transmitted and received signals are related by

$$x = \sqrt{p} s H + v, \quad (2)$$

where  $x \in C^{1 \times N}$  and  $x$  denotes the vector of complex received signals during any given channel use,  $p$  represents the signal-to-noise ratio ("SNR") at the receiver,  $s \in C^{1 \times M}$  and  $s$  denotes the vector of complex transmitted signals,  $H \in C^{M \times N}$  and  $H$  denotes the channel matrix, and the additive noise  $v \in C^{1 \times N}$  and  $v$  is assumed to have independent  $CN(0, 1)$  (zero-mean, unit-variance, complex-Gaussian) entries that are temporally white.

[0135] The channel is used in blocks with there being  $M$  channel uses per block. We can then aggregate the transmit row vectors  $s$  over these blocks of  $M$  channel uses into an  $M \times M$  matrix  $S_\tau$ , where  $\tau = 0, 1, \dots$  represents each block of channel uses. In this setting, the  $m^{\text{th}}$  column of  $S_\tau$  denotes what is transmitted on the  $m^{\text{th}}$  antenna in each instance of time unit  $t$  within block  $\tau$ , and the  $m^{\text{th}}$  row denotes what is transmitted during the  $m^{\text{th}}$  time unit of block  $\tau$  on by each of the  $M$  antennas. If we assume that the channel is constant over the  $M$  channel uses, i.e., over each block  $\tau$ , the input and output row vectors are related through a common channel so that we may represent the received matrix  $X_\tau$  as a function of the transmitted matrix  $S_\tau$  according to the following equation:

$$X_\tau = \sqrt{p}S_\tau H + W_\tau, \quad (3)$$

where  $W_\tau$  and  $H$  are  $M \times N$  matrices of independent  $CN(0,1)$  random variables,  $X_\tau$  is the  $M \times N$  received complex signal matrix and  $S_\tau$  is the  $M \times N$  transmitted complex signal matrix.

[0136] In differential unitary space-time modulation, the transmitted matrix at block  $\tau$  satisfies the following so-called fundamental transmission equation

$$S_\tau = V_{Z_\tau} S_{\tau-1} \quad (4)$$

where  $Z_\tau \in \{0, \dots, L-1\}$  is the data to be transmitted in the form of the code matrix  $V_{Z_\tau}$  (assume  $S_0 = I$ ). Since the channel is used a total of  $M$  times, the corresponding transmission rate is  $R = (Q/M) \log_2 L$ . If we further assume that the propagation environment is approximately constant for  $2M$  consecutive channel uses, then we may write

$$\begin{aligned} X_\tau &= \sqrt{p}S_\tau H + W_\tau = \sqrt{p}V_{Z_\tau} S_{\tau-1} H + W_\tau \\ &= V_{Z_\tau} (X_{\tau-1} - W_{\tau-1}) + W_\tau \end{aligned} \quad (5)$$

which leads us to the fundamental differential receiver equation

$$X_\tau = V_{Z_\tau} X_{\tau-1} + \underbrace{W_\tau - V_{Z_\tau} W_{\tau-1}}_{W_\tau}. \quad (6)$$

Note that the channel matrix  $H$  does not appear in Eq. No. (6), i.e.,  $H$  was substituted for in Eq. No. (5). This implies that, as long as the channel is approximately constant for  $2M$  channel uses, differential transmission permits decoding without knowing the fading matrix  $H$ .

[0137] From Eq. No. (5) it is apparent that the matrices  $V_\ell$  should be unitary, otherwise the product  $S_r = V_{z_r} V_{z_{r-1}} \dots V_{z_1}$  can go to zero, infinity, or both (in different spatial and temporal directions).

[0138] In general, the number of unitary  $M \times M$  matrices in  $\mathcal{V}$  can be quite large. For example, if rate  $R = 8$  is desired with  $M = 4$  transmit antennas (even larger rates are quite possible as shown later), then the number of matrices is  $L = 2^{RM} = 2^{32} \approx 4 \times 10^9$ , and the pairwise error between any two signals can be very small. This huge number of signals calls into question the feasibility of computing a figure of merit,  $\zeta$ , where

$$\zeta = \frac{1}{2} \min_{\ell \neq \ell'} |\det(V_\ell - V_{\ell'})|^{\frac{1}{M}}$$

and lessens its usefulness as a performance criterion. We therefore consider a different, though related, criterion (to be discussed below).

[0139] The large number of signals imposes a large computational load when decoding via an exhaustive search. For high rates, it is possible to construct a random set with some structure. But, again, we have no efficient decoding method. To design sets that are huge, effective, and yet still simple, so that they can be decoded in real-time, it is explained below how the Cayley transform can be used to parameterize unitary matrices.

## 2. The Stiefel Manifold

[0140] The space of  $M \times M$  complex unitary matrices is referred to as the Stiefel manifold. This is the space from which will be selected a subset of unitary matrices that data (to be transmitted) will be mapped-to and received data mapped-from. The Stiefel manifold is highly nonlinear and nonconvex, and can be parameterized by  $M^2$  real free parameters.

## 3. The Cayley Transform

[0141] The Cayley transform of a complex  $M \times M$  matrix  $Y$  is defined to be

$$(I_M + Y)^{-1}(I_M - Y), \quad (7)$$

where  $I_M$  is the  $M \times M$  identity matrix and  $Y$  is assumed to have no eigenvalues at -1 so that the inverse exists (hereafter the  $M$  subscript on  $I$  will be dropped). Note that  $I-Y$ ,  $I+Y$ ,  $(I-Y)^{-1}$  and  $(I+Y)^{-1}$  all commute so there are other equivalent ways to write this transform.



[0142] Compared with other parameterizations of unitary matrices, the parameterization with the Cayley transform is not “too nonlinear” (to be discussed below) and it is one-to-one and easily invertible. The Cayley transform also maps the complicated Stiefel manifold of unitary matrices to the space of skew-Hermitian matrices. Skew-Hermitian matrices are easy to characterize since they form a linear vector space over the real numbers (the real linear combination of any number of skew-Hermitian matrices is skew-Hermitian). And this handy feature will be used below for easy encoding and decoding.

[0143] It should be recognized that the Cayley transform is the matrix generalization of the scalar transform

$$v = \frac{1 - ia}{1 + ia}$$

that maps the real line to the unit circle. This map is also called a bilinear map and is often used in complex analysis. The Cayley transform (Eq. No. 7) maps matrices with eigenvalues inside the unit circle to matrices with eigenvalues in the right-half-plane.

[0144] An important observation to be made is that of Full Diversity. A set of unitary matrices  $\{V_0, \dots, V_L\}$  is fully-diverse, i.e.,  $|\det(V - V_{\ell'})|$  is nonzero for all  $\ell \neq \ell'$  if and only if the set of its skew-Hermitian Cayley transforms  $\{Y_0, \dots, Y_L\}$  is fully-diverse. Moreover, we have

$$V_{\ell} - V_{\ell'} = 2(I + Y_{\ell})^{-1}[Y_{\ell'} - Y_{\ell}](I + Y_{\ell'})^{-1} . \quad (8)$$

[0145] Thus, to design a fully-diverse set of unitary matrices into which will be mapped the data that is to be transmitted, we can design a fully-diverse set of skew-Hermitian matrices and then employ the Cayley transform. This design technique is used in an example below.

#### 4. Cayley Differential Codes

[0146] Because the Cayley transform maps the nonlinear Stiefel manifold to the linear space of skew-Hermitian matrices (and vice-versa) it is convenient to do two things: (1) encode data onto a skew-Hermitian matrix; and then (2) apply the Cayley transform to get a unitary matrix. It is most straightforward to encode the data linearly.

[0147] Again, a Cayley Differential (“CD”) code,  $A$ , is a type of unitary matrix that satisfies the Cayley transform

$$V = (I + iA)^{-1}(I - iA) , \quad (\text{again, } 2)$$

where the CD code, i.e., matrix  $A$ , is Hermitian and is given by

$$A = \sum_{q=1}^Q \alpha_q A_q, \quad (9)$$

and where  $\alpha_1, \dots, \alpha_Q$  are real scalars (chosen from a set  $A$  with  $r$  possible values) (in other words, mapped based upon the data to be transmitted) and where  $A_q$  are fixed  $M \times M$  complex Hermitian matrices.

**[0148]** The code, i.e., the set of unitary matrices  $\{V\}$ , is completely determined by the set of matrices  $A_1, \dots, A_Q$ , which can be thought of as Hermitian basis matrices. Each individual codeword, i.e., each unitary matrix  $V$ , on the other hand, is determined by our choice of the scalars  $\alpha_1, \dots, \alpha_Q$ . Since each  $\alpha_q$  may each take on  $r$  possible values (i.e., the set  $A$  from which values for  $\alpha_q$  are taken has  $r$  values), and the code occupies  $M$  channel uses, then the transmission rate,  $R$ , is  $R = (Q/M) \log_2 r$ . Finally, since an arbitrary  $M \times M$  Hermitian matrix is parameterized by  $M^2$  real variables, we have the constraint

$$Q \leq M^2 \quad (10)$$

Below, as a consequence of the preferred decoding algorithm, a more stringent constraint on  $Q$  will be suggested.

**[0149]** The discussion of how to choose  $Q$  and design the  $A_q$ 's and the set  $A$  follows after the discussion of how to decode  $\alpha_1, \dots, \alpha_Q$  at the receiver.

## 5. Decoding the CD codes

**[0150]** An important property of the CD codes is the ease with which the receiver may form a system of linear equations in the variables  $\{\alpha_q\}$ . To see this, it is useful to substitute the Cayley transform (Eq. No. (1)) into the fundamental receiver equation (Eq. No. (6)),

$$\begin{aligned} X_\tau &= V_{Z_\tau} X_{\tau-1} + W_\tau - V_{Z_\tau} W_{\tau-1} \\ &= (I + iA)^{-1} (I - iA) X_{\tau-1} + W_\tau - (I + iA)^{-1} (I - iA) W_{\tau-1} \end{aligned}$$

implying that

$$(I + iA) X_\tau = (I - iA) X_{\tau-1} + (I - iA) W_\tau - (I - iA) W_{\tau-1},$$

or

$$X_\tau - X_{\tau-1} = A \frac{1}{i} (X_\tau + X_{\tau-1}) + (I + iA) W_\tau - (I - iA) W_{\tau-1}, \quad (11)$$

which is linear in  $A$ . Since the data  $\{\alpha_q\}$  is also linear in  $A$ , then (Eq. No. 11) is linear in  $\{\alpha_q\}$ .

[0151] Consider first the maximum-likelihood estimation of the  $\{\alpha_q\}$ . Using (Eq. No. 17) and noting that the additive noise  $(I+iA)W_T - (I-iA)W_{T-1}$  has independent columns with covariance  $2(I+iA)(I-iA) = 2(I+A^2)$  shows that the maximum likelihood ("ML") decoder is

$$\hat{\alpha}_{ml} = \arg \min_{\{\alpha_q\}} \left\| (I + iA)^{-1} \left( X_\tau - X_{\tau-1} - \frac{1}{i} A (X_\tau + X_{\tau-1}) \right) \right\|^2 ;$$

or, more explicitly,

$$\hat{\alpha}_{ml} = \arg \min_{\{\alpha_q\}} \left\| \left( I + i \sum_{q=1}^Q \alpha_q A_q \right)^{-1} \left( X_\tau - X_{\tau-1} - \frac{1}{i} \sum_{q=1}^Q \alpha_q A_q (X_\tau + X_{\tau-1}) \right) \right\|^2 . \quad (12)$$

[0152] This decoder is not quadratic in  $\{\alpha_q\}$  and so may be difficult to solve. However, if we ignore the covariance of the additive noise in (Eq. No. 11) and assume that the noise is simply spatially white, then we obtain the linearized ML decoder

$$\hat{\alpha}_{lm} = \arg \min_{\{\alpha_q\}} \left\| X_\tau - X_{\tau-1} - \frac{1}{i} \sum_{q=1}^Q \alpha_q A_q (X_\tau + X_{\tau-1}) \right\|^2 \quad (13)$$

We call the decoder "linearized" because the system of equations obtained in solving (Eq. No. 13) for unconstrained  $\{\alpha_q\}$  is linear.

[0153] Because (Eq. No. 13) is quadratic in  $\{\alpha_q\}$ , a simple approximate solution for  $\{\alpha_q\}$  chosen from a fixed set can use nulling and canceling. An exact solution without an exhaustive search can use sphere decoding.

## 6. The number of independent equations

[0154] Nulling and canceling explicitly requires that the number of equations be at least as large as the number of unknowns. Sphere decoding does not have this hard constraint, but it benefits from more equations because the computational complexity grows exponentially in the difference between the number of unknowns and equations. If this difference is not very large, sphere decoding is still feasible.

## 7. Design of the CD Codes

[0155] Although we have introduced the CD code

$$A = \sum_{q=1}^Q \alpha_q A_q, \quad (\text{again, } 9)$$

we have not yet specified  $Q$ , nor have we explained how to design the Hermitian basis matrices  $A_1, \dots, A_Q$  or choose the discrete set  $A$  from which the  $\alpha_q$  are selected. We now address these issues.

### 7.1. Choice of $Q$

[0156] To make the set of all possible CD codes as rich as possible, we should make the number of degrees of freedom  $Q$  as large as possible. Though the parameter  $Q$  can be any size, it has been found that restricting  $Q$  as follows,

$$Q \leq K(2M - K); \quad \text{where } K = \min(M, N), \quad (14)$$

makes it possible to strike a good balance between information content and performance versus computational load. Therefore, as a general practice, we can take  $Q$  at its upper limit in (Eq. No. 14),

$$Q = \min(N, M) \cdot \max(2M - N, M). \quad (15)$$

If sphere decoding is used we sometimes exceed this limit (yielding more unknowns than equations; see examples in Section 3), but  $Q \leq M^2$  should be obeyed.

[0157] We are left with how to design  $A_1, \dots, A_Q$  and how to choose the discrete set  $A$ . If the rates being considered are reasonably small (for example,  $R < 4$ ), then the criterion of maximizing  $|\det(V_{\ell'} - V_{\ell})|$  for all  $\ell' \neq \ell$  is tractable. Recall that any set  $V_1, \dots, V_Q$  for which this determinant is nonzero for all  $\ell' \neq \ell$  is said to be fully diverse.

[0158] It has been shown above (in the discussion of (Eq. No. 8)) that a set of unitary matrices is fully diverse if and only if the corresponding Cayley-transformed set of skew-Hermitian matrices is fully-diverse. Since

$$A' - A = \sum_{q=1}^Q A_q (\alpha'_q - \alpha_q),$$

by considering  $\alpha$  and  $\alpha'$  that differ in only one coordinate  $q$ , we see that it is necessary (but not sufficient) for  $A_1, \dots, A_Q$  to be nonsingular. Some examples of full diversity for small rates and small number of antennas are shown below.

[0159] At high rates, however, it is preferred not to pursue the full-diversity criterion. The reasons include: first, the criterion becomes intractable because of the number of matrices involved; and second, the performance of the set may not be governed

so much by its worst-case pairwise  $|\det(V_\ell - V_{\ell'})|$ , but rather by how well the matrices are distributed throughout the space of unitary matrices. One reason why group sets do not perform very well at high rates is because they lack the required statistical structure of a good high rate set.

## 7.2. Choice of A

[0160] At high rates, a CD code  $A = \sum_{q=1}^Q A_q \alpha_q$  should resemble samples from a Cauchy random matrix distribution. We look first at the implications of the example where there is one transmit antenna  $M = 1$ . In this case the optimal strategy is standard DPSK.

[0161] For the example of  $M = 1$ , we are limited by (Eq. No. 14) to  $Q = 1$  in this example, and there is no loss of generality in setting  $A_1 = 1$  to get

$$v = \frac{1 - i\alpha_1}{1 + i\alpha_1}, \quad \alpha_1 = -i \frac{1 - v}{1 + v} \quad (16)$$

[0162] To get rate  $R = (Q/M) \log_2 L$  with  $M = Q = 1$  we need  $A$  to have  $r = 2^R$  points. Standard DPSK puts these points uniformly around the unit circle at angular intervals of  $2\pi/r$  with the first point at angle  $\pi/r$ . (The location of the first point does not affect the set's performance in any way, but it helps us avoid a formal singularity in the inversion formula (Eq. No. 16) at  $v = -1$ ). For a point at angle  $\Theta$  on the unit circle,

$$\alpha = -i \frac{1 - e^{i\theta}}{1 - e^{-i\theta}} = -\tan(\theta/2). \quad (17)$$

[0163] For example, for  $r = 2$  (D-BPSK, i.e., differential binary PSK), we have  $V = \{e^{\pi i/2}, e^{-\pi i/2}\}$ . Plugging these values into (Eq. No. 17) yields  $\{-1, 1\}$ . For  $r = 4$ , (D-QPSK, i.e., differential quad PSK), we have  $\{-1 - \sqrt{2}, 1 - \sqrt{2}, -1 + \sqrt{2}, 1 + \sqrt{2}\} = \{-2.4142; -0.4142; 0.4142; 2.4142\}$  (where the points are arranged in increasing order). For  $r = 8$ ,  $\{-5.0273, -1.4966, -0.6682, -0.1989, 0.1989, 0.6682, 1.4966, 5.0273\}$ .

[0164] We see that the points rapidly spread themselves out as  $r$  increases, thus reflecting the long tail of the Cauchy distribution  $(p(a) = 1/\pi(1+a^2))$ . We denote  $A$  to be the image of the function (Eq. No. 17) applied to the set  $\theta \in \{\pi/r, 3\pi/r, 5\pi/r, \dots, (2r-1)\pi/r\}$ . In the limit as  $r \rightarrow \infty$ , the fraction of points in  $A$  less than some  $x$  is given by the cumulative Cauchy distribution evaluated at  $x$ .

The  $A$  thus be regarded as an  $r$ -point discretization of a scalar Cauchy random variable.

[0165] While this argument tells us how to choose the set  $A$  as a function of  $r$  when  $Q = M = 1$ , it does not directly show us how to choose  $A$  when  $M > 1$ . Thus, the  $\{\alpha_q\}$  are chosen as discretized scalar Cauchy random variables for any  $Q$  and  $M$ . To complete the code construction,  $\{A_1, \dots, A_Q\}$  should be chosen appropriately, and we present a criterion in the next section.

### 7.3. Choice of $\{A_q\}$

[0166] We shift our attention away from the final distribution on  $A$  and express our design criterion in terms of  $V$ . For a given  $A_1, \dots, A_Q$  and  $A$ , we define a distance criterion for the resulting set of matrices  $V$  to be

$$\xi(V) = \frac{1}{M} E \log \det(V - V')(V - V')^* = \frac{2}{M} E \log |\det(V - V')|, \quad (18)$$

where  $V' = (I + iA')^{-1}(I - iA')$ ,  $A' = \sum_{q=1}^Q A_q \alpha'_q$ , and the expectation is over  $\alpha_1, \dots, \alpha_Q$  and  $\alpha'_1, \dots, \alpha'_Q$  chosen uniformly from  $A$  such that  $\alpha \neq \alpha'$ . Although  $\xi(V)$  is often negative, it is a measure of the expected “distance” between the random matrices  $V$  and  $V'$ .

[0167] To choose the  $A_q$ 's, we therefore propose the optimization problem

$$\arg \max_{A_q = A_q^*, q=1, \dots, Q} \xi(V). \quad (19)$$

[0168] Our choices of  $A_q$  and  $A$  affect the distance criterion through the distribution  $p_V(\bullet)$  that they impose on the  $V$  matrices. It will now be shown that this criterion is maximized when  $V$  and  $V'$  are independently chosen isotropic matrices. Such maximization can be done via gradient-ascent techniques, e.g., Optimization: Theory and Practice, Gordon Beveridge and Robert Shechter, McGraw-Hill, 1970.

[0169] We interpret (Eq. No. 18) as a measure of the average distance between matrices in the set. If the set  $A$  and  $A_1, \dots, A_Q$  are chosen such that  $V$  (namely, the Cayley transform (Eq. No. 1)) is approximately isotropically distributed when  $A$  is sampled uniformly, then the average distance should be large.

[0170] We use (Eq. No. 8), and the fact that matrices commute inside the determinant function, to write the optimization as a function of  $A$  and  $A'$ ,

$$\arg \max_{A_q = A_q^*, q=1, \dots, Q} \left[ \begin{array}{c} \log 4 - \frac{1}{M} E \log \det(I + A^2) \\ - \frac{1}{M} E \log \det(I + A'^2) \\ + \frac{1}{M} E \log \det(A - A')^2 \end{array} \right] \quad (20)$$

where  $A = \sum_{q=1}^Q A_q \alpha_q$ ,  $A' = \sum_{q=1}^Q A_q \alpha'_q$ . For a set with  $\alpha_1, \dots, \alpha_Q$  and  $\alpha'_1, \dots, \alpha'_Q$  chosen from  $A_r$ , we interpret the expectation as uniform over  $A$  such that  $\alpha \neq \alpha'$ .

[0171] It is occasionally useful, especially when  $r$  is large, to replace the discrete set from which  $\alpha_q$  and  $\alpha'_q$  are chosen ( $A$ ) with independent scalar Cauchy distributions. In this case, since the sum of two independent Cauchy random variables is scaled-Cauchy, our criterion simplifies to

$$\arg \max_{A_q = A_q^*, q=1, \dots, Q} \left[ \begin{array}{c} 2 \log 4 - \frac{2}{M} E \log \det(I + A^2) \\ + \frac{1}{M} E \log \det A^2 \end{array} \right] \quad (21)$$

where  $A = \sum_{q=1}^Q A_q \alpha_q$  and the expectation is over  $\alpha_1, \dots, \alpha_Q$  chosen independently from a Cauchy distribution.

#### 7.4. CD Code

[0172] We now summarize the design method for a CD code with  $M$  transmit and  $N$  receive antennas, and target rate  $R$ .

[0173] (i) Choose  $Q \leq \min(N, M) \cdot \max(2M - N, M)$ . This inequality is a hard limit for decoding by nulling/canceling (e.g., G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronic Letters*, Vol. 35, pp. 14–16, Jan. 1999, or G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *J. Sel. Area Comm.*, vol. 17, pp. 1841–1852, Nov. 1999) and  $Q$  is typically chosen to make it an equality. But the inequality is a soft limit for sphere decoding (e.g., U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Mathematics of Computation*, vol. 44, pp. 463–471, April 1985, or M. O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice code decoder for space-time codes,"

IEEE Comm. Let., pp. 161–163, May 2000) and we may choose  $Q$  as large as  $M^2$  even if  $N < M$ .

[0174] (ii) Since  $R = (Q/M) \cdot \log_2(r)$ , set  $r = 2^{MR/Q}$ . Let  $A$  be the  $r$ -point discretization of the scalar Cauchy distribution obtained as the image of the function (Eq. No. 17) applied to the set  $\{\theta\} \sqsubset \{\pi/r, 3\pi/r, 5\pi/r, \dots, (2r-1)\pi/r\}$ .

[0175] (iii) Choose a set  $\{A_q\}$  that solves the optimization problem (Eq. No. 20).

[0176] The following is to be noted.

[0177] A. The solution to (Eq. No. 20) is highly nonunique: simply reordering the  $\{A_q\}$  gives another solution, as does changing the signs of the  $\{A_q\}$ , since the sets  $A$  are symmetric about the origin.

[0178] B. It does not appear that (Eq. No. 20) has a simple closed-form solution for general  $Q$ ,  $M$ , and  $N$ , but presented below is a special case where a closed-form solution appears.

[0179] C. Numerical methods, e.g. gradient-ascent methods, can be used to solve (Eq. No. 20). The computation of the gradient of the criterion in (Eq. No. 20) is presented in the Appendix, the entirety of which is hereby incorporated by reference. Since the criterion function is nonlinear and nonconcave in the design variables  $\{A_q\}$ , there is no guarantee of obtaining a global maximum. However, since the code design can be performed off-line and need only be performed once, one can use more sophisticated optimization techniques that vary the initial condition, use second-order methods, use simulated annealing, etc. Below it is shown that the CD codes obtained with a gradient search tend to have very good performance.

[0180] D. The entries of  $\{A_q\}$  in (Eq. No. 20) are unconstrained other than that the final matrix must be Hermitian. Appealing to symmetry arguments, however, we have found it beneficial to constrain the Frobenius norm of all the matrices in  $\{A_q\}$  to be the same. It is preferred, both for the criterion function (Eq. No. 20) and for the ultimate set performance, that the correct Frobenius norm of the basis matrices be chosen. With the correct Frobenius norm, choosing an initial condition for the  $\{A_q\}$  in the gradient search becomes easier.

[0181] The gradient for the Frobenius norm has a simple closed form which we now give. It can be used to solve for the optimal norm. Let  $\sqrt{\gamma}$  be a multiplicative factor that we use to multiply every  $A_q$ ; we solve for the optimal  $\gamma > 0$  by maximizing the criterion function

$$\arg \max_{A_q = A_q^*, q=1, \dots, Q} \xi(V),$$



that is

$$\begin{aligned} \arg \max_{\gamma} & \left[ 2 \log 4 - \frac{2}{M} E \log \det(I + \gamma A^2) \right. \\ & \left. + \frac{1}{M} E \log \det \gamma A^2 \right] \\ &= \arg \max_{\gamma} \left[ \log \gamma + \frac{2}{M} E \log \det(I + \gamma A^2) \right] \end{aligned}$$

The optimal  $\gamma$  therefore sets the gradient of this last equation to zero:

$$\begin{aligned} 0 &= \frac{1}{\gamma} - \frac{2}{M} E \operatorname{tr} \left[ (I + \gamma A^2)^{-1} A^2 \right] \\ &= \frac{1}{\gamma} - \frac{2}{M} E \operatorname{tr} \left[ ((I + \gamma A^2)^{-1}) \frac{1}{\gamma} (I + \gamma A^2 - I) \right] \\ &= \frac{1}{\gamma} \left( 1 - \frac{2}{M} E \operatorname{tr} \left[ I - (I + \gamma A^2)^{-1} \right] \right) \\ &= \frac{1}{\gamma} \left( -1 + \frac{2}{M} E \operatorname{tr} (I + \gamma A^2)^{-1} \right) \end{aligned}$$

The equation  $-1 + \frac{2}{M} E \operatorname{tr} (I + \gamma A^2)^{-1} = 0$  can readily be solved numerically for  $\gamma$ .

**[0182]** E. The ultimate rate of the code depends on the number of signals sent, namely  $Q$ , and the size of the set  $A$  from which  $\alpha_1, \dots, \alpha_Q$  are chosen. The code rate in bits/channel is

$$R = \frac{Q}{M} \log_2 r \quad . \quad (22)$$

We generally choose  $r$  to be a power of two.

**[0183]** F. The design criterion (Eq. No. 20) depends explicitly on the number of receive antennas  $N$  through the choice of  $Q$ . Hence, the optimal codes, for a given  $M$ , are different for different  $N$ .

**[0184]** G. The variable  $Q$  is essentially also a design variable. The CD code performance is generally best when  $Q$  is chosen as large as possible. For example, a code with a given  $Q$  and  $r$  is likely to perform better than another code of the same rate that is obtained by halving  $Q$  and squaring  $r$ . Nevertheless, it is sometimes advantageous to choose a small  $Q$  to design a code of a specific rate.

[0185] H. If  $r$  is chosen a power of two, a standard gray-code assignment of bits to the symbols of the set  $A$  may be used.

[0186] I. The dispersion matrices  $\{A_q\}$  are Hermitian and, in general, complex.

## 8. Examples of CD Codes and Performance

[0187] Example simulations for the performance of CD codes for various numbers of antennas and rates follow. The channel is assumed to be quasi-static, where the fading matrix between the transmitter and receiver is constant (but unknown) between two successive channel uses. Two error events of interest include block errors, which correspond to errors in decoding the  $M \times M$  matrices  $V_1, \dots, V_L$ , and bit errors, which correspond to errors in decoding  $\alpha_1, \dots, \alpha_Q$ . The bits to be transmitted are mapped to  $\alpha_q$  with a gray code and therefore a block error will correspond to only a few bit errors. In some examples, we compare the performance of linearized likelihood (sphere decoding) with true maximum likelihood and nulling/cancelling.

### 8.1. Simple example: $M = 2$ , $R = 1$

[0188] For  $M = 2$  transmit antennas and rate  $R = 1$ , the set has  $L = 4$  elements. In this case, it turns out that no set can have  $\zeta$ , defined as

$$\zeta = \frac{1}{2} \min_{\ell \neq \ell'} |\det(V_\ell - V_{\ell'})|^{\frac{1}{M}},$$

larger than  $\zeta = \sqrt{2/3} = 0.8165$ . The optimal set corresponds to a tetrahedron whose corners lie on the surface of a three-dimensional unit sphere, and one representation of it is given by the four unitary matrices

$$\begin{aligned} V_1 &= \begin{bmatrix} \sqrt{1/3} + i\sqrt{2/3} & 0 \\ 0 & \sqrt{1/3} - i\sqrt{2/3} \end{bmatrix}, & V_2 &= \begin{bmatrix} -\sqrt{1/3} & \sqrt{2/3} \\ -\sqrt{2/3} & -\sqrt{1/3} \end{bmatrix}, \\ V_3 &= \begin{bmatrix} -\sqrt{1/3} & -\sqrt{2/3} \\ \sqrt{2/3} & -\sqrt{1/3} \end{bmatrix}, & V_4 &= \begin{bmatrix} \sqrt{1/3} - i\sqrt{2/3} & 0 \\ 0 & \sqrt{1/3} + i\sqrt{2/3} \end{bmatrix} \end{aligned} \quad (23)$$

There are many equivalent representations, but it turns out that this particular choice can be constructed as a CD code with  $Q = r = 2$ , and the basis matrices are

$$A_1 = \begin{bmatrix} \frac{1}{\sqrt{2}(\sqrt{3}+1)} & \frac{-i}{\sqrt{2}(\sqrt{3}-1)} \\ \frac{i}{\sqrt{2}(\sqrt{3}-1)} & \frac{-1}{\sqrt{2}(\sqrt{3}+1)} \end{bmatrix}$$

and

$$A_2 = \begin{bmatrix} \frac{1}{\sqrt{2}(\sqrt{3}+1)} & \frac{i}{\sqrt{2}(\sqrt{3}-1)} \\ \frac{-i}{\sqrt{2}(\sqrt{3}-1)} & \frac{-1}{\sqrt{2}(\sqrt{3}+1)} \end{bmatrix}. \quad (24)$$

The matrices (Eq. No. 23) are generated as the Cayley transform (Eq. No. 1) of  $A = A_1\alpha_1 + A_2\alpha_2$ , with  $\alpha_1, \alpha_2 \in \{-1, 1\}$ .

[0189] For comparison, we may consider the set based on orthogonal designs for  $M = 2$  and  $R = 1$  given by

$$V_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad V_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix},$$

$$V_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad V_4 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \quad (25)$$

which has  $\zeta = 1/\sqrt{2} \approx 0.7071$ , or the set

$$V_\ell = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{2\pi i/4} & 0 \\ 0 & e^{2\pi i/4} \end{bmatrix}^{\ell-1}, \quad \ell = 1, \dots, 4$$

which also has  $\zeta = 0.7071$ . Since we are more interested in high rate examples, we do not plot the performance of the CD code (Eq. No. 23); however, simulations show that the performance gain over (Eq. No. 25) is approximately 0.75 dB at high signal to noise ratio ("SNR"). This small example shows that there are good codes within the CD structure at low rates. In this case, the best  $R = 1$  code has a CD structure.

### 8.2. CD Code using orthogonal designs: $M = 2$

[0190] Recall from Lemma 3 that a set of unitary matrices is fully-diverse if and only if its Cayley transform set of skew-Hermitian matrices is fully-diverse. For  $M = 2$  transmit antennas, a famous fully-diverse set is the orthogonal design of Alamouti, namely

$$OD = \begin{bmatrix} x & y \\ -y^* & x^* \end{bmatrix} \quad (26)$$

Orthogonal designs are readily seen to be fully-diverse since

$$\begin{aligned}\det(OD_1 - OD_2) &= \det \begin{bmatrix} x_1 - x_2 & y_1 - y_2 \\ -(y_1 - y_2)^* & (x_1 - x_2)^* \end{bmatrix} \\ &= |x_1 - x_2|^2 + |y_1 - y_2|^2.\end{aligned}$$

We require that OD be skew-Hermitian, implying that

$$OD = \begin{bmatrix} i\alpha_1 & \alpha_2 + i\alpha_3 \\ -\alpha_2 + i\alpha_3 & -i\alpha_1 \end{bmatrix} = i \underbrace{\begin{bmatrix} \alpha_1 & -i\alpha_2 + \alpha_3 \\ i\alpha_2 + \alpha_3 & -\alpha_1 \end{bmatrix}}_{=A} \quad (27)$$

where the  $\alpha_q$ 's are real. Thus, we may define a CD code with basis matrices

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (28)$$

that generates a fully-diverse set. It can be noted that  $A_1, A_2$ , and  $A_3$  form a basis for the real Lie algebra  $su(2)$  of traceless Hermitian matrices. Using (Eq. No. 8) yields

$$\det(V - V') = 4 \det(I + iA)^{-1} \det(A' - A) \det(I + A')^{-1},$$

which upon simplification yields

$$\det(V - V') = \frac{4(|\alpha_1 - \alpha'_1|^2 + |\alpha_2 - \alpha'_2|^2 + |\alpha_3 - \alpha'_3|^2)}{(1 + |\alpha_1|^2 + |\alpha_2|^2 + |\alpha_3|^2)(1 + |\alpha'_1|^2 + |\alpha'_2|^2 + |\alpha'_3|^2)}. \quad (29)$$

For example, by choosing  $\alpha_q \in \{\alpha\}_r$ , we get a code with rate  $R = 1.5$ . The appropriate scaling  $\gamma$  is  $= 1/3$ . The resulting set of eight matrices (which we omit for brevity and because they are readily derived) has  $\zeta = 1/\sqrt{3}$ .

**[0191]** It is noted that the code (Eq. No. 28) is a closed-form solution to (Eq. No. 20) for  $M = 2$  and  $Q = 3$  because it is a local maximum to the criterion.

### 8.3. CD Code vs. OD: $M = N = 2$

**[0192]** For a higher-rate example, we examine another code for  $M = 2$ , but we choose  $N = 2$  and  $R = 6$ . Figure 1 shows the performance of a CD code with  $Q = 4$ . The code is

$$\begin{aligned}A_1 &= \begin{bmatrix} 0.1785 & 0.0510 + 0.1340i \\ 0.0510 - 0.1340i & 0.0321 \end{bmatrix}, \\ A_2 &= \begin{bmatrix} -0.1902 & 0.1230 + 0.0495i \\ 0.1230 - 0.0495i & -0.0512 \end{bmatrix}, \\ A_3 &= \begin{bmatrix} -0.2350 & 0.0515 - 0.0139i \\ 0.0515 + 0.0139i & 0.1142 \end{bmatrix}, \\ A_4 &= \begin{bmatrix} 0.0208 & 0.1143 - 0.1532i \\ 0.1143 + 0.1532i & 0.0220 \end{bmatrix}.\end{aligned}$$

In Fig. 1, the solid line is the block error rate ("bler" in the Fig.) for the CD code with sphere decoding, and the dashed line is the differential two-antenna orthogonal design with maximum likelihood decoding. To get  $R = 6$ , choose  $\alpha_1, \dots, \alpha_4 \in A$  where  $r=8$ ; the distance criterion (Eq. No. 18) for this code is  $\xi = -1.46$ . Also included in the figure is the two-antenna differential orthogonal design with the same rate. The CD code obeys the constraint (Eq. No. 14) and therefore can be decoded very quickly using the sphere decoder. A maximum likelihood decoder would have to search over  $2^{RM} = 2^{12} = 4096$  matrices. It is to be noted that Fig. 1 is not intended to illustrate absolute superior performance, rather it illustrates relatively superior performance.

#### 8.4. Comparison with another nongroup code: $M = 4, N = 1, R = 4$

[0193] There are not many performance curves easily available for existing codes for  $M = R = 4$  over an unknown channel, but the publication by A. Shokrollahi, B. Hassibi, B. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," submitted to IEEE Trans. Info. Theory, 2000, <http://mars.bell-labs.com>, has a nongroup code for  $N = 1$  that appears in Table 4 and Figure 9 of that paper. Figure 2 compares it to a CD code with the same parameters.

[0194] For Fig. 2, the CD code has  $Q = 16$ , and achieves  $R = 4$  by choosing  $r = 2$ . The  $4 \times 4$  matrices  $A_1, \dots, A_{16}$  are:

$$A_1 = \begin{bmatrix} 0 - 0.2404i & 0.0004 - 0.0552i & -0.1191 - 0.1226i & -0.1851 + 0.0590i \\ -0.0004 - 0.0552i & 0 + 0.1088i & -0.0254 + 0.0269i & -0.0037 - 0.0552i \\ 0.1191 - 0.1226i & 0.0254 + 0.0269i & 0 - 0.2123i & -0.0070 + 0.0782i \\ 0.1851 + 0.0590i & 0.0037 - 0.0552i & 0.0070 + 0.0782i & 0 + 0.0803i \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0 + 0.0256i & 0.0680 - 0.0044i & 0.0602 - 0.0450i & 0.1834 - 0.0525i \\ -0.0680 - 0.0044i & 0 + 0.0612i & 0.1373 + 0.0840i & -0.0514 - 0.0203i \\ -0.0602 - 0.0450i & -0.1373 + 0.0840i & 0 - 0.0521i & 0.2104 + 0.1243i \\ -0.1834 - 0.0525i & 0.0514 - 0.0203i & -0.2104 + 0.1243i & 0 + 0.0676i \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 0 - 0.1983i & -0.0603 + 0.0351i & 0.0916 + 0.0494i & 0.1784 - 0.0136i \\ 0.0603 + 0.0351i & 0 + 0.0609i & -0.0190 + 0.0868i & -0.1614 - 0.0575i \\ -0.0916 + 0.0494i & 0.0190 + 0.0868i & 0 + 0.2219i & -0.0621 + 0.0777i \\ -0.1784 - 0.0136i & 0.1614 - 0.0575i & 0.0621 + 0.0777i & 0 - 0.0196i \end{bmatrix}$$

$$A_4 = \begin{bmatrix} 0 + 0.0149i & 0.2168 - 0.0384i & -0.0268 + 0.0702i & -0.0569 - 0.0114i \\ -0.2168 - 0.0384i & 0 + 0.0201i & -0.1388 + 0.0965i & 0.1276 + 0.0793i \\ 0.0268 + 0.0702i & 0.1388 + 0.0965i & 0 - 0.0022i & 0.1576 - 0.0771i \\ 0.0569 - 0.0114i & -0.1276 + 0.0793i & -0.1576 - 0.0771i & 0 + 0.0541i \end{bmatrix}$$

$$A_5 = \begin{bmatrix} 0 - 0.1976i & -0.1267 + 0.0316i & -0.0818 - 0.1269i & -0.0751 - 0.1148i \\ 0.1267 + 0.0316i & 0 - 0.0754i & -0.0671 - 0.1447i & 0.1276 - 0.0364i \\ 0.0818 - 0.1269i & 0.0671 - 0.1447i & 0 + 0.0004i & -0.0336 - 0.0754i \\ 0.0751 - 0.1148i & -0.1276 - 0.0364i & 0.0336 - 0.0754i & 0 - 0.1435i \end{bmatrix}$$

$$A_6 = \begin{bmatrix} 0 - 0.0397i & 0.0041 - 0.0431i & 0.0143 + 0.0019i & -0.0985 - 0.1415i \\ -0.0041 - 0.0431i & 0 + 0.0350i & -0.1616 + 0.1164i & -0.0870 + 0.2128i \\ -0.0143 + 0.0019i & 0.1616 + 0.1164i & 0 + 0.0788i & 0.0720 + 0.0829i \\ 0.0985 - 0.1415i & 0.0870 + 0.2128i & -0.0720 + 0.0829i & 0 + 0.0251i \end{bmatrix}$$

$$A_7 = \begin{bmatrix} 0 + 0.1418i & 0.0244 + 0.1003i & 0.0575 - 0.0581i & -0.0472 - 0.0349i \\ -0.0244 + 0.1003i & 0 - 0.1984i & -0.0059 - 0.0304i & -0.0735 - 0.2520i \\ -0.0575 - 0.0581i & 0.0059 - 0.0304i & 0 + 0.1012i & -0.1113 + 0.0231i \\ 0.0472 - 0.0349i & 0.0735 - 0.2520i & 0.1113 + 0.0231i & 0 + 0.0742i \end{bmatrix}$$

$$A_8 = \begin{bmatrix} 0 - 0.0335i & -0.0190 - 0.1533i & 0.0112 - 0.0098i & -0.0781 - 0.1095i \\ 0.0190 - 0.1533i & 0 - 0.0862i & 0.0116 + 0.1090i & -0.1356 - 0.1393i \\ -0.0112 - 0.0098i & -0.0116 + 0.1090i & 0 + 0.0421i & 0.1032 - 0.1622i \\ 0.0781 - 0.1095i & 0.1356 - 0.1393i & -0.1032 - 0.1622i & 0 + 0.1190i \end{bmatrix}$$

$$A_9 = \begin{bmatrix} 0 + 0.1257i & 0.0192 + 0.0658i & 0.0312 - 0.0430i & -0.0170 + 0.0031i \\ -0.0192 + 0.0658i & 0 + 0.3621i & -0.0893 + 0.0286i & -0.0588 + 0.1090i \\ -0.0312 - 0.0430i & 0.0893 + 0.0286i & 0 - 0.1343i & -0.0469 - 0.1561i \\ 0.0170 + 0.0031i & 0.0588 + 0.1090i & 0.0469 - 0.1561i & 0 - 0.0202i \end{bmatrix}$$

$$A_{10} = \begin{bmatrix} 0 - 0.1220i & -0.0534 + 0.0239i & -0.0291 - 0.0339i & 0.0094 - 0.0649i \\ 0.0534 + 0.0239i & 0 - 0.0337i & -0.1484 + 0.0921i & 0.0813 - 0.0134i \\ 0.0291 - 0.0339i & 0.1484 + 0.0921i & 0 + 0.0159i & -0.1116 - 0.1673i \\ -0.0094 - 0.0649i & -0.0813 - 0.0134i & 0.1116 - 0.1673i & 0 + 0.3019i \end{bmatrix}$$

$$A_{11} = \begin{bmatrix} 0 - 0.0848i & -0.0902 + 0.0471i & -0.0578 + 0.0636i & 0.0540 - 0.0904i \\ 0.0902 + 0.0471i & 0 + 0.0301i & 0.1250 + 0.0087i & 0.0597 - 0.0539i \\ 0.0578 + 0.0636i & -0.1250 + 0.0087i & 0 + 0.2837i & -0.0252 + 0.2114i \\ -0.0540 - 0.0904i & -0.0597 - 0.0539i & 0.0252 + 0.2114i & 0 + 0.0332i \end{bmatrix}$$

$$A_{12} = \begin{bmatrix} 0 + 0.0781i & 0.1675 + 0.0064i & -0.0349 - 0.0324i & 0.1939 - 0.0375i \\ -0.1675 + 0.0064i & 0 - 0.1216i & 0.0977 + 0.0318i & -0.0768 - 0.1414i \\ 0.0349 - 0.0324i & -0.0977 + 0.0318i & 0 - 0.1522i & -0.0964 + 0.0526i \\ -0.1939 - 0.0375i & 0.0768 - 0.1414i & 0.0964 + 0.0526i & 0 + 0.0508i \end{bmatrix}$$

$$A_{13} = \begin{bmatrix} 0 - 0.0117i & 0.0628 - 0.0848i & 0.1047 - 0.1017i & 0.0316 + 0.0272i \\ -0.0628 - 0.0848i & 0 - 0.0322i & -0.0848 + 0.0371i & 0.1228 + 0.0154i \\ -0.1047 - 0.1017i & 0.0848 + 0.0371i & 0 - 0.1907i & -0.2330 - 0.0132i \\ -0.0316 + 0.0272i & -0.1228 + 0.0154i & 0.2330 - 0.0132i & 0 - 0.1408i \end{bmatrix}$$

$$A_{14} = \begin{bmatrix} 0 + 0.1587i & -0.0038 - 0.0996i & -0.1055 - 0.1272i & 0.1282 - 0.1698i \\ 0.0038 - 0.0996i & 0 - 0.0848i & -0.0883 - 0.0334i & -0.0280 + 0.1595i \\ 0.1055 - 0.1272i & 0.0883 - 0.0334i & 0 - 0.0480i & 0.0030 - 0.0765i \\ -0.1282 - 0.1698i & 0.0280 + 0.1595i & -0.0030 - 0.0765i & 0 + 0.0256i \end{bmatrix}$$

$$A_{15} = \begin{bmatrix} 0 - 0.0574i & 0.1189 - 0.0981i & -0.0998 - 0.0472i & -0.0315 - 0.0113i \\ -0.1189 - 0.0981i & 0 + 0.0527i & 0.0116 + 0.1028i & 0.1104 - 0.0912i \\ 0.0998 - 0.0472i & -0.0116 + 0.1028i & 0 + 0.2906i & 0.1081 + 0.0020i \\ 0.0315 - 0.0113i & -0.1104 - 0.0912i & -0.1081 + 0.0020i & 0 - 0.1787i \end{bmatrix}$$

$$A_{16} = \begin{bmatrix} 0 - 0.0315i & 0.0136 + 0.0632i & -0.0392 + 0.1217i & -0.2722 + 0.0216i \\ -0.0136 + 0.0632i & 0 + 0.0899i & -0.0040 - 0.0596i & 0.1057 - 0.0382i \\ 0.0392 + 0.1217i & 0.0040 - 0.0596i & 0 + 0.1765i & 0.0531 - 0.0535i \\ 0.2722 + 0.0216i & -0.1057 - 0.0382i & -0.0531 - 0.0535i & 0 + 0.0908i \end{bmatrix}$$

[0195] In Fig. 2,  $\xi = -1.46$ . The nongroup code, which has its origin in a group code, performs better but the difference is very small. Observe that  $Q = M^2 > 2MN - N^2 = 7$  and therefore the inequality (Eq. No. 14) is not satisfied, but it does not matter in this case because the decoding for both codes is true maximum likelihood (rather than sphere decoding or nulling/cancelling). This example is not very practical because maximum likelihood decoding involves a search over  $2^{RM} = 2^{16} = 65,536$  matrices. However, this same CD code is used in the next example where, by increasing the number of receive antennas to  $N = 2$ , we are able to solve the linearized likelihood with sphere decoding.

### 8.5. Linearized vs. Exact ML: $M = 4, N = 2, R = 4$

[0196] By increasing the number of receive antennas in the previous example to  $N = 2$ , we may linearize the likelihood and compare the performance with the true maximum likelihood. Figure 3 shows the results. In Fig. 3, the solid lines are the block error rate ("bler" in the Fig.) and the bit error rates ("ber" in the Fig.) for the CD code with sphere decoding, and the dashed lines are the block/bit error rates with maximum likelihood decoding. The same CD code as in the example of Figure 2 is used.

[0197] In the example of Fig. 3, observe that  $Q > 2MN - N^2 = 12$  and therefore the inequality (Eq. No. 14) is still not obeyed; but because it is almost obeyed, the sphere decoder of the linearized likelihood searches over only  $16 - 12 = 4$  dimensions. With  $r = 2$ , this search is over  $2^4 = 16$  quantities, which is a negligible burden. Compare this burden with the true maximum likelihood (65,536 matrices). Fig. 3 shows that the performance loss for linearizing the likelihood is approximately 1.3 dB at high SNR. While the performance of linearized maximum likelihood is slightly worse than true maximum likelihood, the next figure shows that the performance of nulling/cancelling is much worse than either.

### 8.6 Sphere decoding vs. nulling/cancelling: $M = N = 4, R = 8$

[0198] Figure 4 shows the performance of a CD code for  $M = 4$  transmit and  $N = 4$  receive antennas for rate  $R = 8$  with linearized-likelihood decoding. In Fig. 4, the solid lines are the block and bit error rates for sphere decoding and the dashed lines are for nulling/cancelling. The performance advantage of sphere decoding is dramatic. As in the previous example of Fig. 3,  $Q = 16$ , but to achieve  $R = 8$  we choose  $r = 4$ . Again, the explicit description of  $A_1, \dots, A_{16}$  is omitted for brevity and because they are readily derived;  $\xi = -1.36$ .

[0199] Also plotted in Fig. 4 is a comparison of the same CD code with nulling/cancelling decoding. We see that sphere decoding is significantly better. True maximum likelihood decoding is not realistic in this example because there are  $2^{RM} - 2^{32} \approx 4 \times 10^9$  matrices in the codebook.

### 8.7 High-rate example: $M = 8, N = 12, R = 16$

[0200] Some of the original V-BLAST experiments use eight transmit and twelve receive antennas to transmit more than 20 bits/second/Hz. Figure 5 shows that high



rates with reasonable decoding complexity are also within reach of the CD codes. In Fig. 5, the solid lines are the block and bit error rates for sphere decoding.

[0201] Plotted in Fig. 5 are the block and bit error rates for  $R = 16$ ; here  $Q = 64$  and  $r = 4$ . The CD matrices are again omitted for brevity and because they are readily derived;  $\xi = -1.48$ . We note that because  $M = 8$ , the effective set size of the set of unitary matrices is  $L = 2^{RM} = 2^{128} \approx 3.4 \times 10^{38}$ , yet we may still easily sphere decode the linearized likelihood.

## 9. Recap

[0202] The Cayley differential codes we have introduced do not require channel knowledge at the receiver, are simple to encode and decode, apply to any combination of one or more transmit and one or more receive antennas, and have excellent performance at very high rates. They are designed with a probabilistic criterion: they maximize the expected log-determinant of the difference between matrix pairs.

[0203] The CD codes make use of the Cayley transform that maps the nonlinear Stiefel manifold of unitary matrices to the linear space of skew-Hermitian matrices. The transmitted data is broken into substreams  $\alpha_1, \dots, \alpha_Q$  and then linearly encoded in the Cayley transform domain. We showed that  $\alpha_1, \dots, \alpha_Q$  appear linearly at the receiver and can be decoded by nulling/cancelling or sphere decoding by ignoring the data dependence of the additive noise. Additional channel coding across  $\alpha_1, \dots, \alpha_Q$  or from block to block can be combined with a CD code to lower the error probability even further.

[0204] Finally, we choose  $\alpha_q$ 's from a set  $A$  designed to help make the final  $A$  matrix behave, on average, like a Cauchy random matrix.

[0205] Applicants hereby incorporate by reference the entirety of their internet-published paper, "Cayley Differential Unitary Space-Time Codes", available at <http://mars.bell-labs.com/cm/ms/what/mars/papers/cayley/>.

[0206] The invention may be embodied in other forms without departing from its spirit and essential characteristics. The described embodiments are to be considered only non-limiting examples of the invention. The scope of the invention is to be measured by the appended claims. All changes which come within the meaning and equivalency of the claims are to be embraced within their scope.

## APPENDIX

## Gradient of criterion (Eq. No. 20)

[0207] In all the simulations presented in this paper the maximization of the design criterion function in (Eq. No. 20), needed to design the CD codes, is performed using a simple constrained-gradient-ascent method. In this section, we compute the gradient of (Eq. No. 20) that this method requires. More sophisticated optimization techniques that we do not consider, such as Newton-Raphson, scoring, and interior-point methods, can also use this gradient.

[0208] From (Eq. No. 20), the criterion function is

$$\begin{aligned} \frac{1}{M} E \log \det(V - V')(V - V')^* &= \log 4 - \frac{1}{M} E \log \det(I + A^2) - \frac{1}{M} E \log \det(I + A'^2) \\ &\quad + \frac{1}{M} E \log \det(A' - A)^2 \end{aligned}$$

where  $A = \sum_{q=1}^Q A_q \alpha_q$  and  $A' = \sum_{q=1}^Q A_q \alpha'_q$ . We are interested in the gradient of this function with respect to the matrices  $A_1, \dots, A_Q$ . To compute the gradient of a real function  $f(A_q)$  with respect to the entries of the Hermitian matrix  $A_q$ , we use the formulas

$$\left[ \frac{\partial f(A_q)}{\partial \operatorname{Re} A_q} \right]_{j,k} = \lim_{\delta \rightarrow 0} \frac{1}{\delta} [f(A_q + \delta(e_j e_k^T + e_k e_j^T)) - f(A_q)], \quad j \neq k \quad (\text{A.2})$$

$$\left[ \frac{\partial f(A_q)}{\partial \operatorname{Im} A_q} \right]_{j,k} = \lim_{\delta \rightarrow 0} \frac{1}{\delta} [f(A_q + i\delta(e_j e_k^T + e_k e_j^T)) - f(A_q)], \quad j \neq k \quad (\text{A.3})$$

$$\left[ \frac{\partial f(A_q)}{\partial A_q} \right]_{j,j} = \lim_{\delta \rightarrow 0} \frac{1}{\delta} [f(A_q + \delta e_j e_j^T) - f(A_q)], \quad (\text{A.4})$$

where  $e_j$  is the  $M$ -dimensional unit column vector with a one in the  $j^{\text{th}}$  entry and zeros elsewhere.

[0209] To apply (Eq. No. A.2) to the second term in (Eq. No. A.1), we compute

$$\begin{aligned} &\log \det(I + (A + (e_j e_k^T + e_k e_j^T) \alpha_q \delta)^2) \\ &= \log \det(I + A^2 + [A(e_j e_k^T + e_k e_j^T) + (e_j e_k^T + e_k e_j^T) A] \alpha_q \delta + O(\delta^2)) \\ &= \log \det[(I + A^2)(I + (I + A^2)^{-1}[A(e_j e_k^T + e_k e_j^T) + (e_j e_k^T + e_k e_j^T) A] \alpha_q \delta + O(\delta^2))] \\ &= \log \det(I + A^2) + \operatorname{tr} \log(I + (I + A^2)^{-1}[A(e_j e_k^T + e_k e_j^T) + (e_j e_k^T + e_k e_j^T) A] \alpha_q \delta + O(\delta^2)) \\ &= \log \det(I + A^2) + \operatorname{tr}[(I + A^2)^{-1}[A(e_j e_k^T + e_k e_j^T) + (e_j e_k^T + e_k e_j^T) A] \alpha_q \delta] + O(\delta^2) \end{aligned}$$

$$\begin{aligned}
&= \log \det(I + A^2) + \begin{bmatrix} [(I + A^2)^{-1} A]_{k,j} \\ + [(I + A^2)^{-1} A]_{j,k} \\ + [A(I + A^2)^{-1}]_{k,j} \\ + [A(I + A^2)^{-1}]_{j,k} \end{bmatrix} \alpha_q \delta + O(\delta^2) \\
&= \log \det(I + A^2) + 4 \operatorname{Re}[(I + A^2)^{-1} A]_{j,k} \alpha_q \delta + O(\delta^2).
\end{aligned}$$

[0210] The last equality follows because  $(I + A^2)^{-1}$  and  $A$  commute and  $A$  is Hermitian. We may now apply (Eq. No. A.2) to obtain

$$\left[ \frac{\partial E \log \det(I + A^2)}{\partial \operatorname{Re} A_q} \right]_{j,k} = 4E \operatorname{Re}[(I + A^2)^{-1} A]_{j,k} \alpha_q, \quad j \neq k$$

[0211] The gradient with respect to the imaginary components of  $A_q$  is handled in a similar way to obtain

$$\begin{aligned}
&\log \det(I + (A + (e_j e_k^T - e_k e_j^T) \alpha_q i \delta)^2) \\
&= \log \det(I + A^2) + \operatorname{tr}[(I + A^2)^{-1} [A(e_j e_k^T - e_k e_j^T) + (e_j e_k^T - e_k e_j^T) A] \alpha_q i \delta] + O(\delta^2) \\
&= \log \det(I + A^2) + \begin{bmatrix} [(I + A^2)^{-1} A]_{k,j} \\ - [(I + A^2)^{-1} A]_{j,k} \\ + [A(I + A^2)^{-1}]_{k,j} \\ - [A(I + A^2)^{-1}]_{j,k} \end{bmatrix} \alpha_q i \delta + O(\delta^2) \\
&= \log \det(I + A^2) + 4 \operatorname{Im}[(I + A^2)^{-1} A]_{j,k} \alpha_q \delta + O(\delta^2)
\end{aligned}$$

which yields

$$\left[ \frac{\partial E \log \det(I + A^2)}{\partial \operatorname{Im} A_q} \right]_{j,k} = 4E \operatorname{Im}[(I + A^2)^{-1} A]_{j,k} \alpha_q, \quad j \neq k$$

[0212] The gradient with respect to the diagonal elements is

$$\left[ \frac{\partial E \log \det(I + A^2)}{\partial A_q} \right]_{k,k} = 2E [(I + A^2)^{-1} A]_{j,j} \alpha_q.$$

[0213] The third term in (Eq. No. A.1) has the same derivative as the second term.

[0214] For the fourth term, note that  $A' - A = \sum_{q=1}^Q A_q \beta_q$  where  $\beta_q = \alpha'_q - \alpha_q$ .

Therefore,

$$\log \det(A + (e_j e_k^T + e_k e_j^T) \delta \beta_q)^2$$

$$\begin{aligned}
&= \log \det(A(I + A^{-1}(e_j e_k^T + e_k e_j^T)\delta\beta_q))^2 \\
&= \log \det(A(I + A^{-1}(e_j e_k^T + e_k e_j^T)\delta\beta_q)A(I + A^{-1}(e_j e_k^T + e_k e_j^T)\delta\beta_q)) \\
&= \log \det A^2 + 2\text{tr} \log(I + A^{-1}(e_j e_k^T + e_k e_j^T)\delta\beta_q) + O(\delta^2) \\
&= \log \det A^2 + 2\text{tr}[A^{-1}(e_j e_k^T + e_k e_j^T)\delta\beta_q] + O(\delta^2) \\
&= \log \det A^2 + 2([A^{-1}]_{k,j} + [A^{-1}]_{j,k})\delta\beta_q \\
&= \log \det A^2 + 4\text{Re}[A^{-1}]_{j,k}\delta\beta_q.
\end{aligned}$$

Hence,

$$\left[ \frac{\partial E \log \det(A' - A)^2}{\partial \text{Re } A_q} \right]_{j,k} = 4\text{Re}[A^{-1}]_{j,k} \beta_q, \quad j \neq k$$

**[0215]** For brevity, the computation of the derivatives with respect to the imaginary and diagonal components of  $A_q$  is omitted. The results are

$$\left[ \frac{\partial E \log \det(A' - A)^2}{\partial \text{Im } A_q} \right]_{j,k} = 4E \text{Im}[A^{-1}]_{j,k} \beta_q, \quad j \neq k$$

and

$$\left[ \frac{\partial E \log \det(A' - A)^2}{\partial A_q} \right]_{j,j} = 2E[A^{-1}]_{j,j} \beta_q.$$

**<Remainder of Page Intentionally Left Blank>**